DOCUMENT RESUME

ED 247 300 TM 840 482

AUTHOR Engle, Molly
TITLE Factors Affecting the Reduction of Narrative Data.
PUB DATE Apr 84
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Content Analysis; *Data Analysis; Educational Objectives; *Experimenter Characteristics; Narration; *Reliability; Research Methodology
IDENTIFIERS *Data Reduction Methods; *Qualitative Data

ABSTRACT

Narrative data yield rich detail, insight, and information, but the personal and situational characteristics of coders (called value inertia and cognitive limitation biases) can affect data reduction. The effects of coder exposure to expected project outcomes and the level of coder research methodology sophistication were investigated. Coders, considered either sophisticated or naive in research methodology, were exposed to positive, ambiguous, or negative expected project outcomes. The coders were then trained and given 25 open-ended interview response sets to reduce or categorize. The response sets were categorized into previously established positive, negative, and ambiguous statement-type, content-code categories using a written protocol. The reliability of coders was investigated by computing generalizability (reliability) coefficients. Generalizability coefficients of .92 were found, regardless of level of exposure to expected project outcomes. This indicates that coders were able to code the same statements in the same way and is an indication of the coders' ability to reproduce the results. Results of this study also indicate that evaluators should use sophisticated coders for narrative data reduction, given that option. Sophisticated coders appear more resistant to the effect of exposure to expected project outcomes. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247300

FACTORS AFFECTING THE REDUCTION OF NARRATIVE DATA

Molly Engle

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Engle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 840 482

Presented at the American Educational Research Association annual meeting, New Orleans, April, 1984

FACTORS AFFECTING THE REDUCTION OF NARRATIVE DATA

Molly Engle
2140 Dolphin Drive
Aptos, California 95003

ABSTRACT

Narrative data enable evaluators to understand other people's viewpoints without predetermining those viewpoints by using preselected questionnaire categories. Narrative data yield rich detail, insight, and information. However, reducing narrative data into meaningful conclusions is difficult and time consuming, and requires attention, commitment, and skill on the part of trained coders. The personal and situational characteristics of coders (called value inertia and cognitive limitation biases) affect data reduction.

The effects of coder exposure to expected project outcomes and the level of coder research methodology sophistication were investigated. Coders, considered either sophisticated or naive in research methodology, were exposed to positive, ambiguous, or negative expected project outcomes. The coders were then trained and given 25 open-ended interview response sets to reduce or categorize. The response sets were categorized into previously established positive, negative, and ambiguous statement-type, content-code categories using a written protocol.

The reliability of coders was investigated by computing generalizability (reliability) coefficients. Generalizability coefficients $\geq .92$ were found, regardless of level of exposure to expected project outcomes. This indicates that coders were able to code the same statements the same way and is an indication of the coders' ability to reproduce the results.

In addition, results of this study indicate that evaluators should use sophisticated coders for narrative data reduction, given that option. Sophisticated coders appear more resistant to the effect of exposure to expected project outcomes, coding narrative data to be more positive with less variability than naive coders when exposed to positive expected project outcomes.

FACTORS AFFECTING THE REDUCTION OF NARRATIVE DATA

Introduction

Evaluation is a multifaceted process, the purpose of which has stimulated much debate. Many authorities consider evaluation essential for decision making and policy development. Among them are Stufflebeam and associates (1971); Weiss (1972); Scriven (1974); Wergin (1976); Deshler, Jaenicke, and Maynard-Moody (1977); and Guba (1978). However, evaluation has not always been viewed as the basis for decision making and policy development. Initially, evaluation was inseparable from measurement, and together they were enmeshed in the scientific paradigm of inquiry (Cuba & Lincoln, 1981). Even now, Rossi and Wright (1977) present the randomized controlled experiment as the ideal model for evaluating the effectiveness of public policy.

Scriven (1974) indicates that many evaluations involve a quantitative component, central to which are statistical conceptualizations. He also alludes to the other aspect of evaluation with which each evaluator must sometimes deal: the qualitative component. Guba (1978) suggests one alternative to the conventional evaluation methodology, that of naturalistic inquiry. Naturalistic inquiry is comprehensive in scope, is demanding in design, and requires an equally sound set of skills as any needed by researchers using the experimental model.

Hettich (1980) and Cronbach (1982) indicate that the experimental control of conventional methodology is not incompatible with naturalistic inquiry.

Choosing one approach over the other is not a critical evaluation issue. Rather, it is appropriate to impose some control on some aspects of data collection when the treatment plan is naturalistic and to do some naturalistic observation when the treatment plan is elaborately controlled. Structured data and naturalistic recording of events are both valuable. Glasser and Strauss (1967) clearly state that both forms of data are necessary; evaluations with strong quantitative designs should also gather qualitative critiques generated by program participants. Campbell (1975, cited in Cronbach, 1982) advocates evaluations in which qualitative and quantitative activities support each other. While the more experimental research methods foster the collection of quantitative data (tests of knowledge, attitudes, and skills; frequency counts; ratings), naturalistic research methods lend themselves to the collection of qualitative data (interviews, questionnaires, anecdotal records).

Naturalistic methods attempt to eliminate a priori assumptions by dealing with relevant outcomes and the most powerful explanatory variables that emerge as the study progresses (Carey, 1980). By collecting data in the "real world", naturalistic methods attend to observation and description of those real world situations. Naturalistic methods are concerned with understanding, explaining, and predicting human behavior in the social milieu. Naturalistic methods are characterized by systematic observation of phenomena. These observations result in descriptive data collected across time and place. Data collection typically involves open-ended interviews, direct observation, or document examination (Smith, 1980).

Data resulting from naturalistic evaluations have many forms. An example is the open-ended narrative, consisting of detailed descriptions of people,

experiences, situations, and events. Qualitative data capture what people (participant-respondents; participant-observers) say in their own words, allowing the data gatherer to understand the world as seen by the participants. Patton (1980) indicates that gathering responses to open-ended questions enables the evaluator to understand and capture other people's viewpoints without predetermining those viewpoints through prior selection of questionnaire categories. The evaluator then provides the framework within which people's responses are accurately and thoroughly represented.

Narrative data collection yields a wealth of detail, insight, and information about the project being evaluated. However, this richness is problematic for the evaluator. Patton (1980), Adams (1982), and Cronbach (1982) indicate that the amount of time required for write-up, coding, and analysis of the volume of notes can be overwhelming. In addition, methods of analysis for narrative data are not well formulated. Although quantified data can offer useful frameworks for organizing and interpreting complex natural behaviors, there is no one single agreed-upon procedure for organizing or reducing the volume of narrative data into meaningful and concise results without losing some of the richness of the narrative report.

One method of data reduction, recommended by Fox (1969), is an intermediate or transitional process called "content analysis." The narrative data are divided into two levels: 1) the manifest level, or what the respondent said with nothing assumed; 2) and the latent level, or the affective meaning or feeling tone of the response. Content analysis includes familiarization with data prior to reduction and is based on various project-specific characteristics. These characteristics form protocols for data reduction which must be specific, clear, and logical in relation to the

program. Since the standardization of data collection fosters objectivity (Cronbach, 1982), the use of established protocols standardize the reduction of data collection. Data reduction requires attention, commitment, and skill on the part of the trained coder who assists in organizing the data. It is the coder whom the evaluator must ultimately trust if the complex information set is to yield clear conclusions and valid interpretations.

The coder brings to the data reduction task abilities and talents unique to that coder. However, because meaningful conclusions about behavior are to be formulated, the coder needs to carefully trained to the task. Ideally, coders would transform the same narrative data into essentially the same final conclusions. In reality, this is not the case. Many writers indicate that coder training and the means by which coder agreement is achieved are vital to narrative data use. Cronbach (1982), among others, indicates that reproducibility is improved by reducing the reliance on individual coder judgement and by using instead an objective guideline. Guba and Lincoln (1981) indicate that personal or situational characteristics affect data reduction; therefore, caution must be taken to guard against potential sources of bias that affect the data reduction process.

Bias in acquiring knowledge, especially about the physical universe, has been a longstanding problem for researchers and evaluators and consequently, also affects coders. Sadler (1981) groups the sources of bias that affect evaluators as well as coders into three broad headings: 1) ethical compromises; 2) value inertias; and 3)cognitive limitations.

Ethical compromises are actions for which the evaluator is personally culpable. Value inertias include the particular background knowledge, prior experience, and world view of the evaluator. Sadler (1981) indicates that the

value inertias "may affect what data are collected and how they are interpreted" (p. 26). Sadler addresses these biases in terms of the evaluator. In many evaluations however, because the evaluator works with data coders to collect and interpret the data, these first two categories also apply to data coders. Both ethical compromises and value inertias are distortions reflecting the individual coder's experience. Elimination of these biases is a necessary but not sufficient condition for an unbiased inquiry.

The third category Sadler discusses is that of cognitive limitations, the limitations in information-processing abilities. Since narrative data reduction occurs with in the individual coder's brain, the inferences drawn from the data can be only as good as the cognitive abilities and mechanisms of the human coder.

Sadler identifies 13 biases affecting the performance of the intuitive data processor (the human coder). Those biases relevant to this study include: 1) data overload, or the amount of data the individual is able to receive, process, and remember; 2) first impressions, or the effects of stimulus order on the judgements made; 3) availability of information, or the ability to retrieve examples and to understand reasons and the display of information; and 4) positive and negative instances, or the tendency to select information that would confirm a hypothesis.

To minimize bias, the evaluator must design an evaluation to minimize these influences in each coding situation. Drawing on the work of Sadler, Shulman and Elstein (1975), and Malitz (1978), this study assessed the effects of several variables upon the way naturalistic narrative data were reduced by coders. The variables explored here included one coder attribute variable and

one manipulated cognitive limitation variable. The attribute variable was the coder's knowledge and sophistication of research methodology. This variable was divided into two categories (naive and sophisticated), on the basis of self-report along a continuum ranging from no knowledge to much knowledge. The variable that dealt with cognitive limitations was the nature of the coder's knowledge of the expected project outcomes; that is whether that knowledge was positively, ambiguously or negatively characterized. It was expected that coder research methodology sophistication and the nature of coder knowledge of project outcomes affects the narrative data reduction.

Methodology

The design for this study addresses two major factors. Factor A was the nature of the coder's exposure to expected project outcomes. Coders were randomly assigned into three levels of Factor A: 1) Group A_1 (positive) was exposed to positive expected project outcome information; 2) Group A_2 (ambiguous) was exposed to expected project outcome information of a mixed nature; 3) Group A_3 (negative) was exposed to negative expected project outcome information. This information was included as part of coder training.

Factor B, an attribute variable, was the coder's research methodology sophistication. This factor had two levels: 1) no research methodology exposure or little research methodology exposure, or "naive"; and 2) some research methodology exposure or a lot of research methodology exposure, or "sophisticated."

The dependent variables were several scores derived from the reduction of narrative data. As it was unlikely that the general-area content codes identified in the protocols were of equal significance, it was clear that their importance needed to be identified. This importance was classified with

a weighting system in which the different possible responses were each assigned a weight indicating its level of importance.

To determine the level of importance, a list of content codes was submitted to a panel of project-knowledgeable experts. A continuum was described in detail, along which the experts were to weight the possible responses. Each judge was asked to independently assign a "value of importance" to each positive, negative and ambiguous statement-type in each content code category.

The panel produced several weights for each type of statement in each content code category which, with good interjudge agreement, should not differ in terms of overall location on the response continuum. The mean of these weights became the weight for the statement type in each content code category. This weight was assigned by the researcher each time that response type occurred. After each response (statement) was coded and weighted, the score for each response set or case was computed using each of several different Indexes.

There were several alternative indexes from which the evaluator could choose for computing the dependent variable. Since this study was concerned with the effect of exposure to expected project outcomes and with the level of research sophistication on the reduction of narrative data and not with the effect of any single Index, several dependent variable Indexes were used. Altogether, five different Indexes were identified as appropriate for calculating the dependent variable scores, only four were used for analysis, as the fifth Index was redundant in light of Indexes THREE and FOUR. It was expected that independent variables' effects would be apparent in all indexes selected. If only one Index had been used, it would not be clear whether the

identified independent variables were producing the effect or if the Index used produced the effect; the outcome picture presented would in this case be limited. The several Indexes used to calculate the dependent variable scores were as follows:

1) Index ONE resulted from subtracting the sum of the weighted statements coded to be negative for a case from the sum of the weighted statements coded to be positive for a case and dividing that value by the total number of statements in the case [(WP- WN)/n];

2) Index TWO resulted from subtracting the total number of statements coded to be negative in a case from the total number of statements coded to be positive in a case and dividing by the total number of statements in a case [(#P-#N)/n];

3) Index THREE resulted from dividing the total number of statements coded to be positive by the total number of statements in a case [#P/n];

4) Index FOUR resulted from dividing the total number of statements coded to be negative in a case by the total number of statements in a case [#N/n];

5) Index FIVE resulted from dividing the total number of statements coded to be ambiguous in a case by the total number of statements in a case [#A/n].

The total number of statements in each case was used as the denominator for computing a case score because cases varied in total number of statements. Use of the average meant that case scores were not influenced by the varying number of statements in a case. Each score reflects the average weight of statements in a case, not the weight of the case.

Index ONE and Index TWO could both result in a negative score. To eliminate negative values from the data sets, a constant of 100 was added to scores calculated with Index ONE and a constant of one was added to scores calculated with Index TWO. The ranges for each Index differed. With all Indexes except Index FOUR, high scores indicated more statements coded to be positive; lower scores indicated more statements coded to be negative.

The project participation responses (the stimulus material) were considered a "subjects" (cases) variable and were crossed with each Factor ABC combination. The coders constituted Factor C, which was nested within each combination of Factors A and B (C:AB). This design was considered a three-factor nested repeated measures or mixed design (Winer, 1971; Myers, 1979).

The narrative data used in this study were collected in 1981 from open-ended interviews of 44 adult staff participants in a team-based nutrition-physical fitness program. Twenty-five interview response sets (cases) were selected as representative of a full-range set of potential interview responses. This prevented positive skew from confounding the data classification and increased the probability of a main effect for Factor A. Each of the 25 response sets (cases) were transcribed and typed to control for penmanship, and all 25 typed responses were photocopied for each coder.

Eighteen coders were selected from university education classes. Coders were selected from the range of research methodology sophistication based upon self-report. Each Factor B level had nine coders and each coder was randomly assigned to one of the three levels of Factor A.

The coders were given copies of the protocols for data reduction, content codes, coding sheets, and the photocopied response sets or cases. Training sessions were conducted to familiarize coders with the purpose of the program and with the protocols. A practice data set was also supplied. Coders were first instructed in the use of the protocols, content codes, and the coding score sheet using this practice data set. During the training session held for each of the three levels of Factor A, the coders were told the nature of the "expected" project outcomes. Coders were instructed to work independently and to return the coding score sheets and the verbatim response transcripts within 21 calendar days.

This study was considered an exploratory rather than a confirmatory study (Cronbach, 1982). Since exploratory work is designed to lead ultimately to more well-defined scientific propositions, it is especially desirable in exploratory research to guard against Type II error. At the same time, however, some protection must be established against Type I error. This concern for balance, given the exploratory nature of this study, led to establishing the Type I level at .10. The GENOVA generalized analysis of variance computer program (Brennan, 1983) was used to analyze the results.

Results and Conclusions

The two major factors addressed in this study were the nature of coder exposure to knowledge of expected project outcomes based on three different condensed versions of field reports (Factor A) and levels of coder research methodology sophistication (Factor B). The study concerned the effect of those two factors on narrative data reduction, in this study defined as classification, or coding, done by coders.

The dependent variables were several scores derived from the classification of 25 different interview response sets, or cases, using each of several Indexes. Each coder coded 25 different cases and each case was scored on four Indexes.

Effectiveness of Coder Training

If, in the training sessions, coders were accurately instructed in the rationale, purpose and use of the catetgory system, then coders should have been able to code the same statements the same way. Coding the same statements in the same way would mean that coders are highly reliable.

The GENOVA generalized analysis of variance program was used to determine interjudge reliability. This program produces a coefficient of

generalizability, which is the ratio of universe-score variance to observed-score variance (Allen & Yen, 1979). The reliability coefficient is its counterpart in classical true-score theory. Table 1 reports reliability coefficients (generalizability coefficients) for naive (B_1) and sophisticated (B_2) coders separately for each level of Factor A and for each level of Factor A separately by dependent variable Index.

Reliability (generalizability) coefficient for each level of exposure to project outcome expectations (Factor A) were $\geq$ .92 on all Indexes. Reliability coefficients for both naive and sophisticated coders (Factor B), regardless of level of exposure to project outcome expectations, were $\geq$ .86 on all Indexes. The high reliability coefficients indicate sufficient power to perform a test of significance on the main effects.

Interaction and Main Effects Using Multiple Indexes

In addition to a main effect A (exposure) and main effect B (research methodology), an effect for coders (C:AB) and an effect for cases (S) were computed. Interaction effects between A and B, between A and S, between B and S, and among A, B, and S were also computed.

Results of Analysis. The results reported in Table 2 present the F statistics and the probability levels for each source of variance for Indexes ONE through FOUR. Quasi F ratio statistics were calculated for each main Factor effect and for the interaction of the main Factors.

Table 2 reports no probability level less than .25 for Factor A (exposure) and only Index FOUR was significant at the .10 level for Factor B (research methodology). Coders, represented by C:AB, and cases, represented by S, were both significant at the .01 level for all Indexes.

The AS interaction was significant for Index ONE ($p \leq .10$) and Index FOUR

TABLE 1

RELIABILITY COEFFICIENTS FOR EACH LEVEL OF FACTOR B
WITHIN EACH FACTOR A LEVEL AND FOR EACH LEVEL OF FACTOR A

FACTOR	INDEX 1	INDEX 2	INDEX 3	INDEX 4
A_1 B_1	.9813	.9806	.9899	.9617
A_1 B_2	.9432	.9404	.9415	.9106
A_1	.9788	.9767	.9821	.9693
A_2 B_1	.9066	.9108	.9146	.8707
A_2 B_2	.9683	.9693	.9644	.8673
A_2	.9683	.9702	.9713	.9336
A_3 B_1	.8868	.9199	.9181	.8690
A_3 B_2	.9732	.9769	.9855	.9014
A_3	.9579	.9669	.9747	.9221

TABLE 2

F VALUES AND PROBABILITY LEVELS BY DEPENDENT VARIABLE INDEX FOR EACH SOURCE OF VARIANCE

Index	ONE		TWO		THREE		FOUR	
Effect	F	p	F	p	F	p	F	p
A	1.07*	>.25	0.05*	>.25	0.26*	>.25	1.31*	>.25
B	2.76*	≤.25	1.43*	>.25	0.26*	>.25	2.26*	≤.10
C:AB	3.56	≤.01	5.70	≤.01	5.72	≤.01	3.99	≤.01
S	102.69	≤.01	116.79	≤.01	125.99	≤.01	52.80	≤.01
AB	2.10*	≤.25	3.35*	≤.10	1.43*	>.25	5.78*	<.05
AS	1.43	≤.10	1.29	≤.25	1.18	≤.25	1.47	≤.05
BS	1.23	≤.25	1.28	≤.25	0.76	>.25	1.07	>.25
ABS	1.61	≤.05	1.83	≤.01	1.24	≤.25	1.78	≤.01

*Quasi-F ratio

($p \leq .05$). The BS interaction was nonsignificant for all Indexes. The ABS interaction was significant for Index ONE ($p \leq .05$), Index TWO ($p \leq .01$), and Index FOUR ($p \leq .01$).

Discussion. Since coders are assumed to differ because of inherent coder characteristics, the results of Factor C were expected. This is also true with cases, as it is known that cases differ. The significance of cases (S) contributed to the AS, BS, and the ABS interactions at probability levels equal to or less than .25. This occurs with all Indexes and all interactions except the BS interaction for both Index THREE and Index FOUR, which have a probability level greater than .25. Because the interactions with (S) are unimportant to this study, no post hoc analyses were done.

Results of AB Interaction Analysis. The AB interaction was significant for Index TWO ($p \leq .10$) and Index FOUR ($p \leq .05$). In addition, the AB interaction for Index ONE resulted in a probability of $\leq .25$. A multiple comparison test, Tukey's HSD (Winer, 1971), was applied to all pairwise comparisons among AB means. Significant differences were found between the A_1B_1 and A_3B_1 and between the means of A_3B_1 and A_3B_2.

Cell means for each AB cell combination by Index are presented in Table 3; standard deviations are reported in Table 4. Coder means for each cell combination by Index are presented in Table 5.

Discussion of AB Interaction. Interaction occurs when the effects of independent variables act together on the dependent variable to produce a separate effect. With Indexes ONE and FOUR demonstrating an AB interaction ($p \leq .10$), and Index TWO demonstrating a trend ($p \leq .25$), it may be concluded that the independent variables acted together to produce a separate effect on the dependent variable. Examination of AB cell combination means (Table 3)

TABLE 3

AB CELL MEANS BY DEPENDENT VARIABLE INDEX

	INDEX ONE			INDEX TWO	
	B_1	B_2		B_1	B_2
A_1	94.1	98.9	A_1	83.8	98.8
A_2	107.4	96.3	A_2	103.7	89.9
A_3	107.8	96.7	A_3	109.0	86.7

	INDEX THREE			INDEX FOUR	
	B_1	B_2		B_1	B_2
A_1	38.4	43.8	A_1	54.7	45.6
A_2	46.7	39.9	A_2	42.9	48.8
A_3	42.4	39.4	A_3	32.1	54.0

	INDEX FIVE	
	B_1	B_2
A_1	6.9	10.6
A_2	10.4	11.3
A_3	25.5	6.6

TABLE 4

AB CELL STANDARD DEVIATIONS BY DEPENDENT VARIABLE INDEX

	INDEX ONE			INDEX TWO	
	B_1	B_2		B_1	B_2
A_1	49.96	50.57	A_1	72.60	72.56
A_2	49.07	47.22	A_2	72.38	69.79
A_3	49.17	48.24	A_3	65.18	68.76

	INDEX THREE			INDEX FOUR	
	B_1	B_2		B_1	B_2
A_1	37.10	36.81	A_1	37.33	37.51
A_2	37.38	35.88	A_2	36.94	36.09
A_3	35.80	35.65	A_3	33.79	35.95

TABLE 5

CODER CELL MEANS BY DEPENDENT VARIABLE INDEX

INDEX ONE
[(ΣWP-ΣWN)/n]

A_1		A_2		A_3	
B_1	B_2	B_1	B_2	B_1	B_2
99.0	98.5	95.2	102.6	104.2	95.2
88.7	99.9	121.6	93.6	117.2	92.6
94.7	98.3	192.5	92.6	101.9	102.4

INDEX TWO
[(#P-#N)/n]

A_1		A_2		A_3	
B_1	B_2	B_1	B_2	B_1	B_2
88.7	98.7	88.5	100.8	100.4	81.0
76.6	93.9	128.2	84.5	125.2	81.5
86.2	103.9	94.5	84.4	101.4	97.4

TABLE 5, Continued

INDEX THREE
(#P/n)

A_1		A_2		A_3	
B_1	B_2	B_1	B_2	B_1	B_2
40.1	46.2	40.7	41.2	36.6	40.5
35.9	40.7	59.4	38.7	53.5	37.7
39.3	44.5	39.9	39.9	37.1	40.1

INDEX FOUR
(#N/n)

A_1		A_2		A_3	
B_1	B_2	B_1	B_2	B_1	B_2
51.5	47.5	52.3	40.4	36.1	59.6
59.3	46.8	31.2	54.3	24.4	58.3
53.2	42.4	45.4	51.6	35.8	44.1

reveals that for each of the Indexes reported, sophisticated coders coded more statements to be positive (as evidenced by higher scores) when exposed to positive expected project outcomes. Naive coders, on the other hand, tended to code more statements to be positive when exposed to ambiguous or negative expected project outcomes. The pattern occurred regardless of which Index was used. Because of the probability level associated with Index FOUR ($p \leq .05$), the clearest picture of this pattern is presented with this Index. However, as Indexes THREE, FOUR, and FIVE are all related, these three Indexes must be considered together in interpreting the interaction. It may be that some characteristic of negative and ambiguous statements influence coders exposed to various expected project outcomes.

The results of multiple comparison tests indicate differences between means A_1B_1 and A_3B_1 and between means A_3B_1 and A_3B_2. The difference between the A_1B_1 and A_3B_1 means indicates that naive coders tended to code more statements to be negative when exposed to positive project outcome expectations than did naive coders when exposed to negative project outcome expectations. The difference between A_3B_1 and A_3B_2 means indicates that naive coders tended to code more statements to be positive when exposed to negative project outcome expectations than did sophisticated coders when exposed to the same expectations.

Discussion of the AB interaction needs to include comparison of Indexes THREE, FOUR, and FIVE (Table 3). In Indexes THREE and FOUR, when naive coders were exposed to positive project outcome expectations, they coded more statements to be negative than did sophisticated coders. When exposed to ambiguous or negative project outcome expectations, naive coders coded more statements to be positive than did sophisticated coders. The conclusion might

be drawn that Indexes THREE and FOUR are each other's complement, as the same pattern of coding is observed. However, when statements were coded to be ambiguous (Index FIVE), a different view appeared and the perception of the interaction altered. Naive coders when exposed to positive project outcome expectations coded fewer statements to be ambiguous than did sophisticated coders. When exposed to ambiguous project outcome expectations, there was little difference between cell means of sophisticated and naive coders. When exposed to negative project outcome expectations, naive coders coded more statements to be ambiguous than did sophisticated coders. When comparing Indexes THREE, FOUR, and FIVE, it appears that there is not just an interaction between Factor A (exposure) and Factor B (research methodology), but a three-way interaction between exposure, research methodology, and statements being coded. This may be the result of how statements are worded as well as the combination of exposure and research methodology.

Although Index THREE (positive index) showed probability levels greater than .25 for the AB interaction, this Index must be included in the discussion as it adds clear support to the conclusion for Index FOUR (negative index), that the negative statements were affected by project outcome expectations. If indeed, Index THREE (positive index) and Index FOUR (negative index) were ipsative, then they would reflect the same results. This is not the case, however. Since a significant interaction was found with Index FOUR (negative index) the negative statements must be affected by some other factor. In this case, levels of exposure to expected project outcomes influence the way statements are coded. The statements coded to be ambiguous appear to be contributing to the significance found with Index FOUR.

The trend identified with examination of these coder cell means tends to

indicate that naive coders are more influenced by exposure to mixed/ambiguous and negative project outcome expectations than sophisticated coders. Sophisticated coders tend to be more influenced by positive project outcome expectations. The same trends are observed with all four Indexes. Index FOUR presents the clearest picture of the effects of the AB interaction.

Discussion

The results of this study indicate that coder performance is affected by a combination of each coder's previous experience and the situation in which the coding is done. These interactive biases affect the outcome of data reduction and, consequently, a project's evaluation. As implied by Sadler, the biases of value inertia and cognitive limitations interact to affect the ultimate outcome of data reduction and, therefore, the project outcomes. The evaluator can limit these effects through the coder selection process.

The results of this study indicate that naive coders react more to exposure to project outcome expectations and tend to code statements more positively than do sophisticated coders. Sophisticated coders, on the other hand, tend to be more restrictive in their statement coding. Since every project has the potential for positive, ambiguous, and negative field reports (which tend to be the basis for various project outcome expectations), it is unreasonable to expect that coders would be immune to the effect of field reports. It is also unreasonable to expect coders to be isolated in order to avoid exposure to continuing field reports, although that certainly would be one way to control for the effect of expected project outcomes generated by field reports.

Consequently, when evaluators select coders for narrative data reduction, the use of sophisticated coders would best minimize the effect of personal and

situational biases, as sophisticated coders demonstrate less variability (Table 3 and Table 4). The less variability, the less the effect of the personal and situational biases. However, evaluators often have little choice in the selection of coders. The evaluator has several alternatives under these conditions. The evaluator can select an even mix of sophisticated and naive coders, training the naive coders on the implications of personal biases. In this situation, it is possible that the coders' results would tend to balance. The evaluator needs to be aware that this is not the ideal situation. Another alternative would be to select only naive coders and, in the process of training them in the use of coding protocols, also train them on the implications of personal biases. A third alternative would be to hire as many sophisticated coders as funds would allow, even though the final total number of available coders may be less than desired.

In any case, coder training in the use of coding protocols becomes a primary consideration. The results of this study indicate that sessions for training coders in the use, rational, and purpose of the category system increases the ability of the coders to code the same statements the same way. This is evidenced by the magnitude of the reliability coefficients (Table 1). It does not, however, give any indication as to what other effect training may have on data reduction. It also does not address the issues implicitly raised with alternatives mentioned above.

Summary

Narrative data reduction completed by trained coders leads to meaningful conclusions about the project. By increasing the reproducibility of results and reducing reliance on individual coder judgements, coder training contributes to minimizing personal and situational biases.

Using sophisticated coders, training naive coders, or using an even mix of sophisticated and trained naive coders will increase the likelihood that narrative data will yield meaningful results. These results are, in turn, a valuable contribution to any evaluation as they yield a wealth of information which may or may not reflect the quantitative data. By capturing what people say in their own words, the evaluator uses narrative data to understand the world as seen by project participants. Qualitative data collected from naturalistic methods, when combined with data from other sources, will provide the evaluator with a comprehensive perspective on a program. When data are collected from qualitative and quantitative methods, each supplements the conclusions drawn from the other, better enabling the evaluator to use these conclusions in decision making and policy development.

REFERENCES

Adams, K. A. The keen-edged feather: Intuitive analysis and reporting in qualitative analysis. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April, 1981.

Agnew, N. M., & Pike, S. W. The science game (3rd ed.). Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1982.

Allen, M. J., & Yen, W. M. Introduction to measurement theory. Monterey, Calif.: Brooks/Cole Publishing Company, 1979.

Azrin, N. H. A strategy for applied research: Learning based but outcome oriented. American Psychologist, 1977, 32(2), 140-149.

Brennan, R. L. Elements of generalizability theory. Iowa City, Iowa: The American College Testing Program, 1983.

Carey, R. F. Empirical vs. naturalistic research? Reading Research Quarterly, 1980, 15(3), 412-415.

Carpenter, J. Rationale for and effects of utilizing both qualitative and quantitative analysis in a naturalistic research project. Paper presented at the Adult Educational Research Conference, Minneapolis, Minn., April, 1977.

Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1977.

Coulter, C. L. Training observers for naturalistic observational research. (Report No. 76-9). Austin, Texas: The University of Texas, 1976. (ERIC Document Reproduction Service No. ED 147 274)

Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30(2), 116-127.

Cronbach, L. J. Designing evaluations of educational and social programs. San Francisco: Jossey-Bass Publishers, 1982.

Deshler, D., Jaenicke, D., & Maynard-Moody, S. A policy-oriented evaluation of educational support services for family day care. Paper presented at the Adult Educational Research Conference, Minneapolis, Minn., April, 1977.

Feldt, L. S. A test of the hypothesis that Cronbach's alpha and Kuder-Richardson coefficient twenty is the same for two tests. Psychometrika, 1969, 34(3), 363-373.

Fox, D. J. The research process in education. New York: Holt, Reinhart and Winston, Inc., 1969.

Glass, G. V. Integrating findings: The meta analysis of research. In L. S. Shulman (Ed.), Review of Research in Education, 1977, 5, 351-379.

Glasser, B. G., & Stauss, A. L. The discovery of grounded theory. Chicago: Aldine, 1967.

Guba, E. G. Toward a methodology of naturalistic inquiry in educational evaluation. Monograph Series, No. 8. Los Angeles: Center for Study of Evaluation, University of California, 1978.

Guba, E. G., & Lincoln, Y. S. Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey-Boss, Publishers, 1981.

Hall, G. E. Ethnographers and ethnographic data: An iceberg of the first order for the research manager. Education and Urban Society, 1980, 12(3), 349-366.

Hammond, R. G. Evaluation at the local level. In B. R. Worthen and J. R. Sanders, Educational evaluation: Theory and practice. Worthington, Ohio: Jones, 1973.

Hettich, P. The evaluators journal: A qualitative adjunct to program evaluation. CEDR Quarterly, 1980, 13(2), 19-22.

Houtkooper, L. C., Engle, M., Simmons, L., Petersen, K., Davidson, M., Tinsley, A., & Gibbs, J. Interim report. Nutrition Super Stars and Arizona Nutrition Education and Training Program, Nutritional Education Region Resource Centers. Arizona Department of Education, AG contract #1587, No. 187, October, 1981.

Kerlinger, F. N. Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart and Winston, Inc., 1973.

Lofland, J. Analyzing social settings: A guide to qualitative observation and analysis. Belmont, Calif.: Wadsworth Publishing Company, Inc., 1971.

Malitz, D. Identification of predictors of coding performance: Selecting subjects to be trained to use observation systems. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March, 1978.

Myers, J. L. Fundamentals of experimental design (3rd ed.). Boston, Mass.: Allyn and Bacon, Inc., 1979.

Patton, M. Q. Qualitative evaluation methods. Beverly Hills, Calif.: Sage Publications, 1980.

Popham, W. J. Educational evaluation. Englewood Cliffs, N. J.: Prentice-Hall, 1975.

Provus, M. Discrepancy evaluation. Berkeley, Calif.: McCutchan, 1971.

Rist, R. C. On relations among educational research paradigms: From disdain to detente. Anthropology and Education Quarterly, 1977, 8(2), 42-49.

Rossi, P. H., & Wright, S. R. Evaluation research: An assessment of theory, practice, and politics. Evaluation Quarterly, 1977, 1, 5-52.

Sadler, D. R. Intuitive data processing as a potential source of bias in naturalistic evaluations. Educational Evaluation and Policy Analysis, 1981, 3(4), 25-31.

Scriven, M. Goal free evaluation. In E. R. House (Ed.), School evaluation: The politics and process. Berkeley, Calif.: McCutchan, 1973.

Scriven, M. Evaluation perspectives and procedures. In W. J. Popham (Ed.), Evaluation in education. Berkeley, Calif.: McCutchan, 1974.

Seltiz, C., Johoda, M., & Cook, S. W. Research methods in social relations (3rd ed.). New York: Holt, Rinehart and Winston, 1978.

Shulman, L. S., & Elstein, A. S. Studies of problem solving, judgement, and decision making: Implications for educational research. In F. N. Kerlinger (Ed.), Review of Research in Education, 1975, 3, 3-42.

Smith, M. L. Naturalistic research, The Personnel and Guidance Journal, 1981, 59(9), 585-589.

Stake, R. E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stake, R. E. (Ed.) Evaluating the arts in education: A responsive approach. Columbus, Ohio: Merrill, 1975.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., Provus, M. M. Educational evaluation and decision making. Itasca, Ill.: Peacock, 1971.

Tikunoff, W. J., & Ward, B. A. Conducting naturalistic research on teaching: Some procedural considerations. Education and Urban Society, 1980, 12(3), 263-290.

Tinsley, A. M., Houtkooper, L. G., Engle, M., & Gibbs, J. A nutrition and physical fitness test for fifth and sixth graders. Journal of Nutrition Education, 1981, 13(3), 95-96.

Tyler, R. W. Basic principles of curriculum and instruction: Syllabus for education 360. Chicago: University of Chicago Press, 1949.

Weiss, C. H. Evaluation Research. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Wergin, J. F. The evaluation of organizational policy making: A political model. Review of Educational Research, 1976, 46(1), 75-115.

Winer, B. J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill Book Company, 1971.

Wolf, R. L., & Tymitz, B. Toward more natural inquiry in education. CEDR Quarterly, 1977, 10(3), 7-9.